Firm Business Networks

Chien-Hung Chien, Armin Haller and Anton H. Westveld

Australian National University
{chien-hung.chien,armin.haller,anton.westveld}@anu.edu.au
http://www.anu.edu.au

Abstract. This paper describes an ontology-based approach to integrate datasets from Intellectual Property Australia and the Australian Securities Exchange to study firms in business networks. We combine different indicator variables with SPARQL queries for research to understand the characteristics of different firms in multiple business networks. We use an exponential random graph models approach to describe factors that help firms form business networks. In doing so we find evidence of homophily for large firms in patents and trademarks business networks. They are more likely to form business networks in comparison with small & medium firms. For firms in patents and shared director and trademarks and directors business networks, firm size does not play an important factor in the formation of business networks and there is limited evidence of homophily.

Keywords: semantic web, business networks, exponential random graph model

1 Introduction

At a time when governments face budgets constraints, it is important for them and their statistical agencies to make better use of available resources. Governments around the world have realised the advantages of integrating their datasets to use them for purposes beyond which they were collected for. The Australian Government's open data agenda aims to integrate multiple data sources and provide information to encourage evidence-based policy development [1]. The 2017 Productivity Commission inquiry into *Data Availability and Use* highlighted the need to create integrated and linked national interest datasets to inform policy development [2].

The need to integrate a large number of datasets from multiple sources has created a big data challenge for statistical offices, including the Australian Bureau of Statistics (ABS). The statistical challenges associated with creating and analysing data from diverse sources has been discussed extensively in [3]. A recent paper by [4] has presented several ABS case studies on using semantic web technologies to visualise and analyse integrated datasets. This preliminary research builds on [5]. Using open and purchased data sources, we focus on combining semantic web and statistical methods to develop a better understanding of firms in multiple business networks.

Firms seek partners with complementary assets to leverage each other's strengths and find competitive advantages to ensure market success. Business networks play a vital role in finding new market opportunities and obtaining the necessary resources to achieve growth [6]. There are different types of business networks ranging from more structured (business groups or franchising) to less structured (R&D consortium, trade association and shared directors). These business networks facilitate different degrees of knowledge transfer and create social capital to enhance business performance [7].

Business networks play a particularly important role in ensuring the economic success of small firms. Firms in business networks have mutual dependence to ensure each other's success. Business networks can also help better resource allocation and reduce operational risks through cooperative arrangements. This is particularly important in sectors with fast technological advancement and short product life cycle. This is evident by the success of high-tech start-ups in Taiwan, where business networks play an important role in integrating the operation of a large number of specialised small firms in subcontracting and outsourcing industries [8, p.2-4].

There is empirical evidence to support firm R&D collaboration as an important source of innovation to improve firm performance [9,10]. R&D collaboration enables knowledge transfer between firms to share new managerial ideas and technology. Firms look for different competitive advantages in the market through business networks [11,12]. Firms can also form business networks through shared directors. [13] argued that boards of directors can enhance firm performance through effectively monitoring and providing resources. [14] also found that directors serve as an important asset to form business networks, particularly for young high-tech firms.

Governments use a range of initiatives to encourage business networks to achieve economic growth. The Australian Government, through the Department of Industry, Innovation and Science, actively supports firm collaboration. It has established Cooperative Research Centres to encourage collaborative research partnerships between research institutes and industry partners [15]. Another example is Innovation Connections which provides funding to support collaborative projects between small and medium sized firms to find expert technology advice and collaborate with research centres in developing new ideas with commercial potential [16].

It is useful for policy makers to have an understanding of the factors that could lead to the formation of business networks. Are the factors contributing to firms forming business networks different when we compare firms in multiple business networks? Are firms with similar characteristics more likely to form business networks? What kind of impact did the global financial crisis (GFC) have on firms in multiple business networks? This study uses open data and purchased data to address these questions.

2   Semantic web

We use an ontology-based data model to integrate three datasets - patents, trademarks, and publicly listed companies - to answer our research questions. The patents and trademarks datasets are from 2017 Intellectual Property Government Open Data (IPGOD). The publicly listed companies on shared directors from the Australian Securities Exchange (ASX) is purchased from MorningStar DatAnalysis Premium. Examples of data visualisation can be found in Appendix B.

The semantic web approach is well suited to integrate data from multiple sources and to extract information on firms in multiple business networks. We assign a unique Uniform Resource Identifier (URI) for each firm using a unique firm identifier e.g. Australian Business Numbers (ABNs) or Australian Company Numbers (ACNs). We attach different firm attributes e.g. Industry, Profit and Loss etc. from different data sources to each firm. Our datasets include Australian firms in three different types of business network relationships: firms that collaborate in R&D related activities (patents networks), firms that collaborate for commercial interests (trademarks networks) and firms that share directors (shared directors networks). We are comparing firms belonging to multiple networks i.e. patents and shared directors networks, patents and trademarks networks and trademarks and shared directors networks. In addition, we separate our sample into three periods: before the global financial crisis (GFC) from 2003 to 2006, during the GFC from 2007 to 2009 and after the GFC from 2010 to 2013 in our sample.

For our analysis, it is important to know the data provenance to correctly compare firms with and without multiple business networks. We use named graphs in the knowledge graph to distinguish different data sources. These named graphs are then used in the SPARQL queries to retrieve the correct subgraph. Figure 1 shows the ontology for our data model. We use Ontodia - an OWL and RDF diagramming tool to visualise our data model. For example, $firm\_alpha$ is connected to $firm\_beta$ within a patent business network, while it is also connected with $firm\_gamma$ in a trademark business network. In comparison, $firm\_beta$ is in a patent business network with $firm\_alpaha$ and shares a director with $firm\_delta$.

We have $518,870$ triples of firms in the patent named graph http://patents, $5,638,915$ triples of firms in the trademark named graph http://trademarks and $272,289$ triples in directors named graph http://directors with a total of $6,430,074$ triples in the integrated knowledge graph. We use legal entities to represent firms. The IPGOD and ASX datasets contain unique firm identification numbers (ABNs or ACNs) for firms. We use patent and trademark applications (application number), and directors (director id) to identify firms in different business networks. We use unique ABNs and ACNs to form the URI for the legal entities. These URIs serve as unique linking keys to correctly retrieve firm information from different sources using SPARQL queries. The bottom right panel of Figure 1 shows the basic business network relationships. The Business Network node qualifies how firms can be connected through joint patent

Fig. 1: Ontology
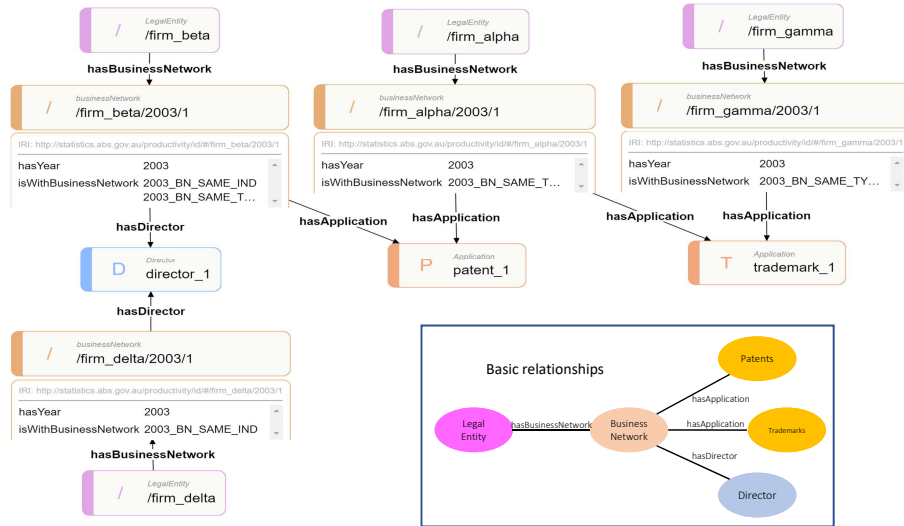
or trademark applications or shared directors. An example below shows how we construct a SPARQL query to retrieve firms belonging to both patent and trademark networks in the period before the GFC from 2003 to 2006.

Listing 1.1: intersection SPARQL query

```
prefix pat: <http://patents>
prefix tmk: <http://trademarks>
SELECT ?ABN
FROM NAMED pat:
FROM NAMED tmk:
WHERE {
values (?BN) {("2003_BN") ("2004_BN") ("2005_BN") ("2006_BN")}
{GRAPH pat:{
?LegalEntity fnet:hasAustralianBusinessNumber ?ABN;
fnet:hasBusinessNetwork ?businessNetwork.}}
FILTER EXISTS
{GRAPH tmk: {
?LegalEntity fnet:hasAustralianBusinessNumber ?ABN;
fnet:hasBusinessNetwork ?businessNetwork.}}}
```

3   Data sources

3.1   Intellectual Property Government Open Data

Patents and trademarks datasets are from the 2017 Intellectual Property Government Open Data (IPGOD) for this study. IPGOD contains administrative

information on patents and trademarks [17]. Patent and trademark applications can be filed by one applicant or multiple applicants. Over the sample period between 2003 and 2013, there are $329,809$ applicant-patent application combinations with $290,568$ unique patent applications. In comparison, there are $1,350,134$ applicant-trademark application combinations with $639,958$ unique trademark applications. Counting unique ABNs and unique ACNs, $15,617$ firms filed patent applications and $152,539$ firms filed trademark applications over the sample period between 2003 and 2013.

We do not observe whether firms are in business networks or not. However, we observe if a firm files a patent or trademark application by itself or with another firm(s) in IPGOD. Therefore, we define a firm as being in a business network in year t when it shares a patent and/or a trademark application with at least one other firm. We create indicator variables equal to one if a firm has a patent and/or trademark application with at lease one other applicant type (small & medium or large enterprises) in year t. The indicator takes zero value if a firm files an application by itself. One would support that the business networks that generate joint patent and/or trademark applications could have existed before year t and could have lasted beyond year t. Consider one scenario, firm A, which had a joint application with firm B in 2003. The last available observation for firm A is in 2005. The network indicator will show firm A was in a business network from 2003 to 2005. We have made this assumption because the duration of a standard patent is 25 years and 10 years for a trademark right [18][19].

3.2 Purchased publicly listed company data

The data on publicly listed companies and their directors is purchased from MorningStar DatAnalysis Premium data service. This service contains detailed reports for all current and formerly listed companies on the Australian Securities Exchange (ASX). There are $1,722$ listed companies and $9,892$ directors in the sample reference period between 2003 and 2013. We use the following data items in our analysis - unique director identification number (DirectorID), ABNs, ACNs, Global Industry Classification Standard (GICS) industry sectors, GICS industry groups, director appointed dates and director resigned dates.

We create an indicator variable if a firm shares a director with at least one other firm during the sample period. If a director's appointed date is before $01-01-2003$, we use $01-01-2003$ as the appointed date. We exclude directors who resigned before $01-01-2003$. The duration of the shared director network is derived by taking into account the director's earliest appointed and latest resigned dates. For example, if director 001 worked in firm A between 2003 and 2004 and firm B between 2004 and 2005 then firm A and B are connected in the director network between 2003 to 2005. There is a higher proportion of firms with network connections in the ASX data compared to IPGOD.

3.3 Multiple business networks

The scope of the analysis includes firms in three types of multiple business networks. They include when a firm files a joint patent and a joint trademark application with at least one other firm, when a firm files a joint patent application and shares the same director with at least one other firm, and when a firm files a joint trademark application and shares the same director with at least one other firm. Figure 2 compares the proportion of large and small & medium enterprises in these multiple business networks. The number of business networks has grown over time in the sample. Overall, there are no significant differences between the proportion of large and small & medium enterprises over different periods. However, we observe a larger increase in the proportion of small & medium enterprises in multiple networks when we compare between the before GFC (2003 to 2006) and during GFC (2007 to 2009) periods.

Fig. 2: Proportion of applicants types between 2003 and 2013 in business networks
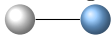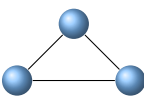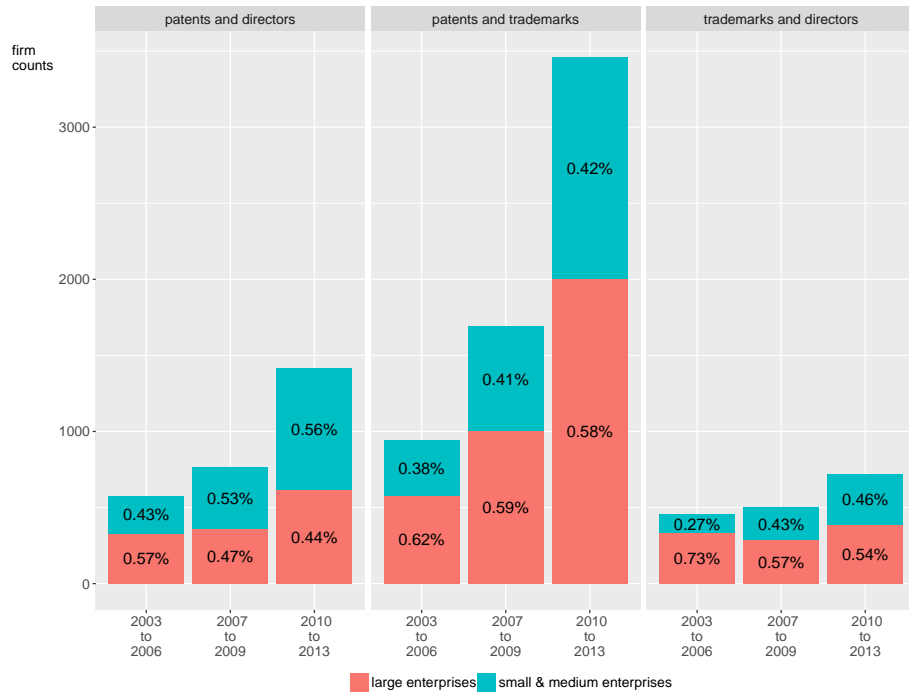
4   Statistical model

Our research goal is to describe the factors that contribute to the formation of business networks. There are many interdependent social processes that drive the formation of business networks. The formation of business networks can be influenced by the presence (or absence) of other ties in the network. The complexity is shown by the business network formed between Verizon Wireless and Google in 2009. Verizon Wireless, one of the key wireless telecommunications providers in the United States, wanted to become less reliant on Apple and iPhone to deliver its service to high-paying customers [20]. This was mainly because AT&T, one of Verizon Wireless's main competitors, had already established a closed working relationship with Apple [21]. The successful partnership between Verizon Wireless and Google has led to forming of business networks with other Android phone manufacturers like Samsung [22].

Business networks are inherently relational so the occurrence of a particular relationship, or tie, could depend on the occurrence of other ties [23]. The above example clearly demonstrates the endogenous network structural effects. Firms consider factors beyond simple evaluation of the suitable characteristics of the perspective partners. The decision by Verizon Wireless involves a strategic response to compete against AT&T by forming other ties with Google and Samsung. An observed business network can result from a combination of simultaneous processes with interdependent endogenous factors [24].

We use an exponential random graph models (ERGMs) approach which takes into account the underlying network structure, characteristics of firms and the characteristics of the dyad in the inference. ERGMs have two main functions: (1) to describe if a given network structure, e.g. edge or transitivity observed in a network, occur more than expected by chance; (2) to determine whether there is an association between network links and firm characteristics and between network links and dyad characteristics or both [25]. Table 1 shows selected commonly observed business network structures [26].

Table 1: Common business network structures

Network structure	Graphic configuration	Network statistics
edge		$\sum_{i,j \in N} X_{ij} X_{ji}$ for dissimilar nodes
homophily		$\sum_{i,j \in N} X_{ij} X_{ji}$ for similar nodes
transitivity		$\sum_{i,j,k \in N} X_{ij} X_{jk} X_{ki}$

Note that the dependency structure is shown when a tie being included in homophily can also be contained in the transitivity structure .

We build on the work of [27] and [28] and use ERGMs to study Australian business networks. We use formula (2) in Appendix C to define ERGMs as

$$Pr(\boldsymbol{Y} = \boldsymbol{y} \mid \boldsymbol{\theta}, \boldsymbol{X}) = \frac{1}{k(\boldsymbol{\theta})} \exp\left[\boldsymbol{\theta}^{\mathsf{T}} g(\boldsymbol{y}, \boldsymbol{X})\right], \tag{1}$$

where $\boldsymbol{y}$ is the observed business network and it takes value 1 if there is a tie between firm i and firm j and 0 otherwise. The symbol $\boldsymbol{\theta}$ represents unknown parameters of interest and determines the effects of the network statistics. We use $g(\boldsymbol{y}, \boldsymbol{X})$ to represent the endogenous network statistics in the model. We follow [28] and include edges and transitivity structures, common network structures observed in business network data. We use $\boldsymbol{X}$ to denote exogenous explanatory variables for the network statistics. There are three firm-specific characteristics. The variable *Products* is the number of patents or trademark applicants registered by a firm. Our measures of homophily are *LargeFirm* and *SME*. The *LargeFirm* indicates a tie is formed between two large enterprises. The *SME* indicates a tie between two small & medium enterprises. The reference group is a tie form between a *LargeFirm* and a *SME*. The dyad-specific characteristic, *Industry*, takes the value one if a network contains at least two firms in the same industry or zero otherwise. The term $k(\boldsymbol{\theta})$ is the normalising constant. A discussion on the key ERGMs assumption and basic concept can be found in Appendix C.

4.1   Results

We use the R `ergm` package - The Statnet Project to estimate our models. See [29] for more details. Some useful references and tutorials can be found in [30,31]. Table 2 shows the statistical model results. The results are in log-odds as discussed in Appendix C. Firms with higher numbers of patent or trademark applications have a slightly higher probability of forming business network. We have some evidence of homophily, i.e. firms with similar characteristics are more likely to form business networks with each other. Table 1 shows that the coefficients for *LargeFirm* and *SME* are significant. The probabilities are generally higher for *LargeFirm* (0.92) than *SME* firms (0.73) in particular for firms in patent and trademark business networks in the period before the GFC. We do not observe homophily for firms in trademark and shared director networks as the coefficients for both *LargeFirm* and *SME* are insignificant in the three periods. There is a mixed result for the same industry indicator variable on firms. Firms operating in same industries appear to have a higher probability of forming trademark and shared director business networks than patent and shared director networks. See model results in Appendix A.

5   Conclusion and future direction

This preliminary analysis has shown the benefits of using semantic web to integrate datasets to study firms in business networks. We have found that large

firms are more likely to form business networks when they are in patents and trademarks business networks. The homophily results are mixed for firms in patents and shared director and trademarks and shared directors business networks. Our research could be extended in several areas. One possibility is to combine the open and purchased datasets with ABS business data to obtain more firm characteristics, such as better industry classifications or firm productivity to improve the statistical models. Another possibility is to compare the model results with latent class model approach to better understand business network effects.

Acknowledgements

References

1. Turnbull, M.: Speech to the locate 15 conference: The power of open data (12 2015) accessed at link on 2017-06-03.
2. Commission, P., et al.: Data availability and use. Inquiry Report **82** (2017) accessed at link on 2017-08-11.
3. Tam, S.M., Clarke, F.: Big data, official statistics and some initiatives by the australian bureau of statistics. International Statistical Review **83**(3) (2015) 436–448
4. Clarke, F., Chien, C.H.: Visualising big data for official statistics: The abs experience. In: Data Visualization and Statistical Literacy for Open and Big Data. IGI Global (2017) 224–252
5. Clarke, F., Chien, C.H.: Connectedness and Meaning: New Analytical Directions for Official. In: Proceedings of the 3rd International Workshop on Semantic Statistics, Bethlehem, U.S. (October 2015) page where your paper starts–ends accessed at link on 2018-04-01.
6. Lee, C., Lee, K., Pennings, J.M.: Internal capabilities, external networks, and performance: A study on technology-based ventures. Strategic Management Journal **22**(6/7) (2001) 615–640
7. Inkpen, A.C., Tsang, E.W.K.: Social capital, networks, and knowledge transfer. The Academy of Management Review **30**(1) (2005) 146–165 accessed at link on 2017-08-03.
8. Perry, M., Knapp, A.B., Piggott, V.C.: Small Firms and Network Economies. Routledge (2002)
9. Ahuja, G.: Collaboration networks, structural holes, and innovation: A longitudinal study. Administrative Science Quarterly **45**(3) (2000) 425–455 accessed at link.
10. Belderbos, R., Carree, M., Lokshin, B.: Cooperative r&d and firm performance. Research Policy **33**(10) (2004) 1477–1492 accessed at link.
11. Miotti, L., Sachwald, F.: Co-operative r&d why and with whom? Research Policy **32**(8) (2003) 1481–1499 accessed at link.

12. Chuluun, T., Prevost, A., Upadhyay, A.: Firm network structure and innovation. Journal of Corporate Finance **44** (2017) 193–214 accessed at link.
13. Hillman, A.J., Dalziel, T.: Boards of Directors and Firm Performance: Integrating Agency and Resource Dependence Perspectives. The Academy of Management Review **28**(3) (2003) 383–396
14. Collins, C.J., Clark, K.D.: Strategic Human Resource Practices, Top Management Team Social Networks, and Firm Performance: The Role of Human Resource Practices in Creating Organizational Competitive Advantage. The Academy of Management Journal **46**(6) (2003) 740–751
15. Department of Industry, Innovation and Science: Cooperative research centres programme (2017) accessed at link on 2018-04-01.
16. Department of Industry, Innovation and Science: Innovation connections (2018) accessed at link on 2018-06-24.
17. IP Australia: Intellectual property government open data 2017 (2017) accessed at link on 2017-08-03.
18. IP Australia: Patent basics (2018) accessed at link on 2018-04-01.
19. IP Australia: Trademark basics (2018) accessed at link on 2018-05-16.
20. Svensson, P.: Verizon, google in android partnership (6 2009) accessed at link on 2018-12-03.
21. Cohan, P.: Project Vogue: Inside Apple's iPhone Deal With AT&T (9 2013) accessed at link on 2018-12-03.
22. Tibken, S.: Samsung, verizon will partner on 5g smartphone in first half of 2019 (12 2018) accessed at link on 2018-12-03.
23. Koskinen, J., Daraganova, G.: Exponential random graph model fundamentals. In Lusher, D., Koskinen, J., Robins, G., eds.: Exponential random graph models for social networks: Theory, methods, and applications. Cambridge University Press (2013) 49–76
24. Kim, J.Y., Howard, M., Cox Pahnke, E., Boeker, W.: Understanding network formation in strategy research: Exponential random graph models. Strategic management journal **37**(1) (2016) 22–44
25. Valente, T.W.: Social networks and health: Models, methods, and applications. Volume 1. Oxford University Press New York (2010)
26. Robbins, G., Dean, L.: Simplified account of an exponential random graph model as a statistical model. In Lusher, D., Koskinen, J., Robins, G., eds.: Exponential random graph models for social networks: Theory, methods, and applications. Cambridge University Press (2013) 29–36
27. Desmarais, B., Cranmer, S.: Statistical mechanics of networks: Estimation and uncertainty. Physica A: Statistical Mechanics and its Applications **391**(4) (2012) 1865 – 1876 accessed at link.
28. Balland, P.A., Belso-Martínez, J.A., Morrison, A.: The dynamics of technical and business knowledge networks in industrial clusters: Embeddedness, status, or proximity? Economic Geography **92**(1) (2016) 35–60
29. Handcock, M.S., Hunter, D.R., Butts, C.T., Goodreau, S.M., Krivitsky, P.N., Morris, M.: ergm: Fit, Simulate and Diagnose Exponential-Family Models for Networks. The Statnet Project (`http://www.statnet.org`). (2018) R package version 3.9.4.
30. Hunter, D.R.: Curved exponential family models for social networks. Social networks **29**(2) (2007) 216–230
31. Levy, M.: Ergm tutorial. = http://michaellevy.name/blog/ERGM-tutorial/ (2016) Accessed: 22-09-2018.

32. Lybbert, T.J., Zolas, N.J.: Getting patents and economic data to speak to each other: an 'algorithmic links with probabilities' approach for joint analyses of patenting and economic activity. Research Policy **43**(3) (2014) 530–542

33. Cranmer, S.J., Desmarais, B.A.: Inferential network analysis with exponential random graph models. Political Analysis **19**(1) (2011) 66–86

34. Cranmer, S.J., Leifeld, P., McClurg, S.D., Rolfe, M.: Navigating the range of statistical tools for inferential network analysis. American Journal of Political Science **61**(1) (2017) 237–251

35. Morris, M., Handcock, M.S., Butts, C.T., Hunter, D.R., Goodreau, S.M., de Moll, S.B., , Krivitsky, P.N.: Exponential random graph models (ergms) using statnet tutorial (2016) accessed at link on 2019-01-03.

36. Goodreau, S.M., Kitts, J.A., Morris, M.: Birds of a feather, or friend of a friend? using exponential random graph models to investigate adolescent social networks. Demography **46**(1) (2009) 103–125

37. Martina Morris, Steven M. Goodreau, S.M.J.: Network modeling for epidemics (2018) accessed at link on 2019-01-03.

Appendix A Model results

Appendix B Visualisation

Fig. 3 in Appendix shows an example of firms in shared directors business networks using information from MorningStar DatAnalysis Premium. The firms are classified to different industries using Global Industry Classification Standard (GICS). GICS Classification provides details of the classification. Each node represents a firm which shares a director with at least one other firm. The colour of the node represents different industries classifications.

Table 2: model comparison

	Dependent variable:								
	Networks before GFC 2003 2006			Networks during GFC 2007 2009			Networks after GFC 2010 2013		
	PAT and ASX	PAT and TMK	TMK and ASX	PAT and ASX	PAT and TMK	TMK and ASX	PAT and ASX	PAT and TMK	TMK and ASX
edges	−4.194***	−10.546***	−6.181***	−4.571***	−11.191***	−6.214***	−5.014***	−11.741***	−6.290***
	(0.571)	(0.329)	(0.497)	(0.571)	(0.287)	(0.414)	(0.518)	(0.257)	(0.317)
triangles	0.939***	4.469***	1.570***	1.017***	5.811***	1.539***	0.983***	8.501***	1.882***
	(0.279)	(0.612)	(0.228)	(0.234)	(0.387)	(0.186)	(0.206)	(0.503)	(0.193)
Products	0.022	0.011	0.029***	0.028	0.032***	0.030***	0.035*	0.051***	0.033***
	(0.026)	(0.013)	(0.008)	(0.020)	(0.009)	(0.007)	(0.018)	(0.011)	(0.006)
Industry	−0.076		0.754***	0.286		0.718***	0.406		0.574***
	(0.270)		(0.241)	(0.284)		(0.203)	(0.264)		(0.154)
Largefirm	0.899**	2.025***	0.278	0.371	1.536***	0.191	0.569**	2.038***	0.088
	(0.406)	(0.438)	(0.192)	(0.301)	(0.399)	(0.169)	(0.275)	(0.331)	(0.140)
SME	1.162***	−1.091**	0.553	0.541*	−1.092***	0.183	0.532*	−1.206***	−0.088
	(0.408)	(0.432)	(0.568)	(0.321)	(0.340)	(0.460)	(0.306)	(0.302)	(0.463)
Akaike Inf. Crit.	221.305	1,151.229	712.345	345.917	3,167.493	1,048.500	441.311	4,756.693	1,259.173
Bayesian Inf. Crit.	247.979	1,210.979	746.731	375.303	3,233.346	1,086.537	473.130	4,827.291	1,298.919

Note: *p<0.1; **p<0.05; ***p<0.01

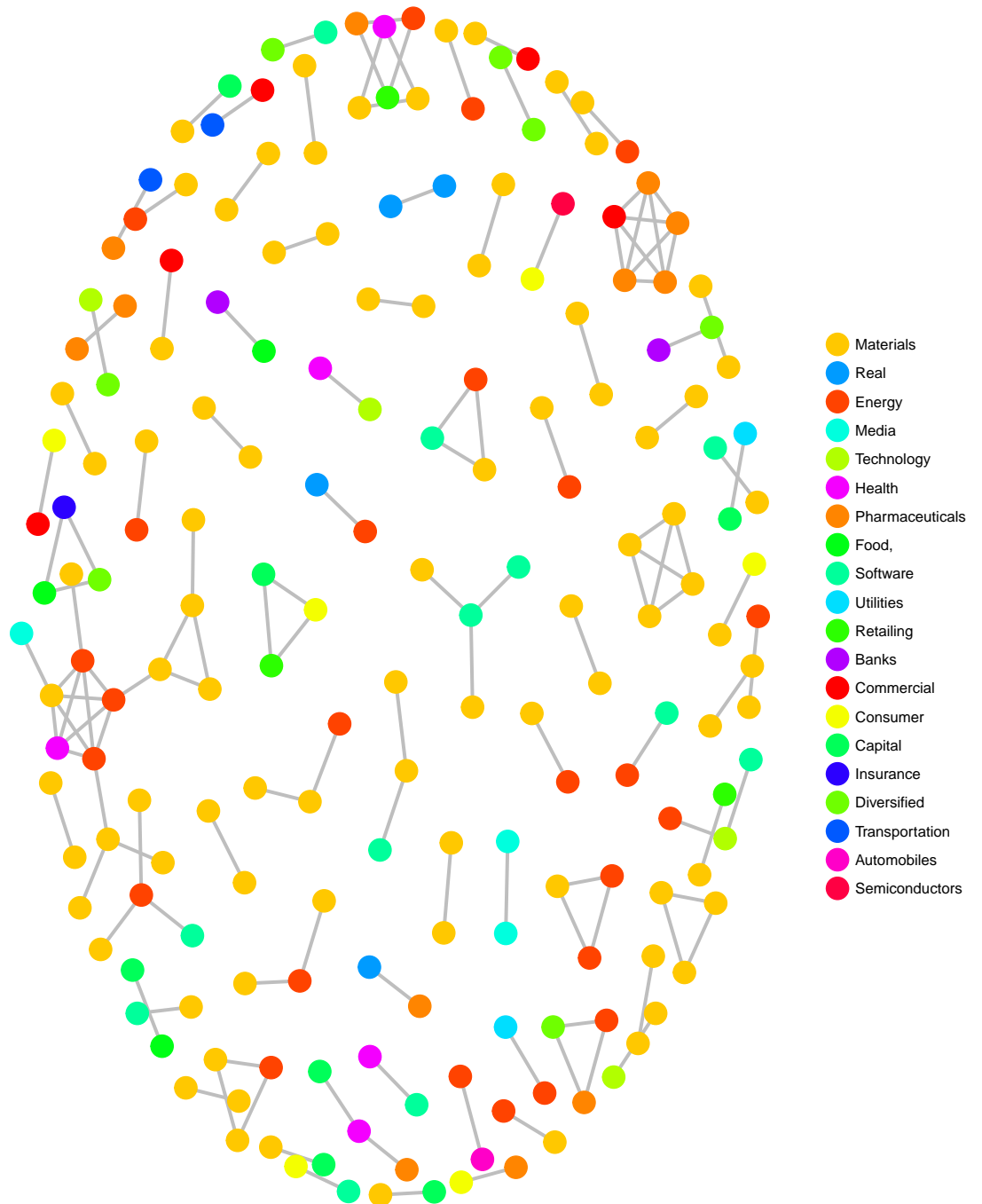Fig. 3: Firm networks through shared directors

Fig. 4 in Appendix shows the innovation hotspots across Australia. We combine the available latitude and longitude coordinates information with the product type information from 2017 Intellectual Property Government Open Data . Each node represent the location of an applicant. There is missing latitude and longitude information. However, these firms have postcode information. We use a combination of information from the Post Office, Google map api and a research dataset to impute latitude and longitude information for these firms. The use of postcode is an improvement than using the capital cities for imputing missing latitude and longitude information.

The colour of the node represents the product type for patent or trademark applications. Patent and trademark applications are classified using different classifications. This makes it difficult to compare innovations between patent and trademark applications. [32] proposed an algorithmic links with probabilities (ALP) approach, which analyses the text descriptions, to concord International Patent Classification (IPC) for patents and Nice Agreement (NICE) Classifications for trademarks. We applied their research and concord trademark applications to IPC.

Fig. 4: Innovation hot spots
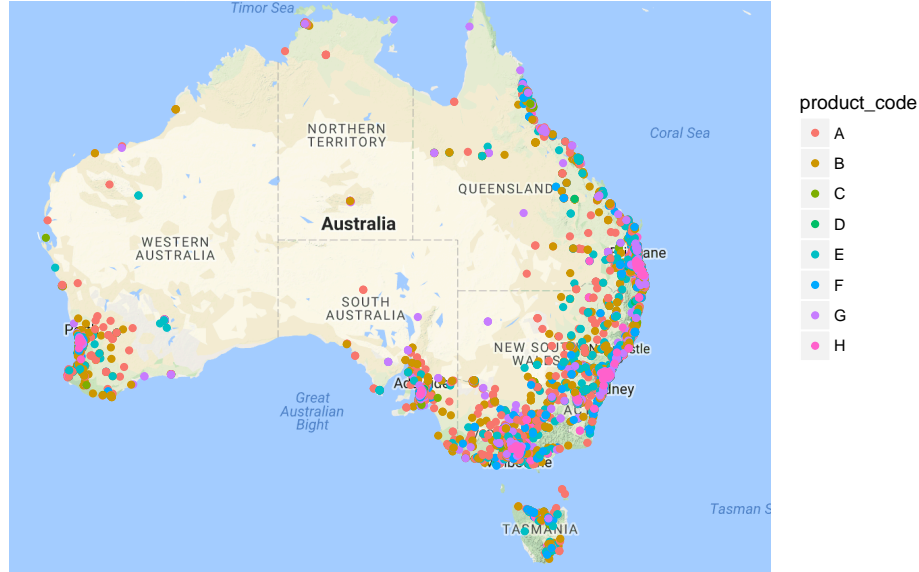
A = Human Necessities, B = Performing Operations, Transporting, C = Chemistry, Metallurgy, D = Textiles and Paper, E = Fixed Constructions, F = Mechanical Engineering, G = Physics, H = Electricity.

Appendix C   ERGMs assumptions and basic concept

The exponential random graph models (ERGMs) maximise the probability of the observed networks over the networks with the same number of vertices that could have been observed to estimate parameters. The approach allows statistical inference without independence assumptions. This is because the approach allows for endogenous dependencies coming from the networks. It also contains a set of network statistics that can include exogenous variables coming from the characteristics of vertices or edges [33,34].

We follow [35] and specify the general form for an ERGM as:

$$P(\boldsymbol{Y} = \boldsymbol{y}) = \frac{1}{k(\boldsymbol{\theta})} \exp[\boldsymbol{\theta}^{\mathsf{T}} g(\boldsymbol{y}, \boldsymbol{X})], \tag{2}$$

where $\boldsymbol{Y}$ is the random variable for the state of the network and $\boldsymbol{y}$ is the observed networks, $g(\boldsymbol{y}, \boldsymbol{X})$. We use $\boldsymbol{X}$ to denote the observed firm characteristics. The symbol $\boldsymbol{\theta}$ represents unknown parameters of interest and determines the effects of the network statistics. The symbol $k(\boldsymbol{\theta})$ is the normalising constant, it represents the quantity in the numerator summed over all possible networks (typically constrained to be all networks with the same number of node set as $\boldsymbol{y}$). The formula (2) can be re-expressed in terms of the conditional log-odds of a single tie between two actors as

$$\mathrm{logit}\,(Y_{ij} = 1 \mid \boldsymbol{y}_{ij}^{\mathsf{C}}) = \boldsymbol{\theta}^{\mathsf{T}} \delta(\boldsymbol{y}_{ij}), \tag{3}$$

where Y_{ij} is the random variable for the state of the firm pair i,j (with realisation $\boldsymbol{y}_{ij}$). We use $\boldsymbol{y}_{ij}^{\mathsf{C}}$ to denote the complement of $\boldsymbol{y}_{ij}$, i.e. all connections in the network except y_{ij}. The vector $\delta(\boldsymbol{y}_{ij})$ contains the change statistic for each model term. The change statistic records how $g(\boldsymbol{y}, \boldsymbol{X})$ term changes when y_{ij} is toggled from 0 to 1 [36]. So

$$\mathrm{logit}\left[Pr(Y_{ij} = 1 \mid \boldsymbol{y}_{ij}^{\mathsf{C}})\right] = \log\left[\frac{(Y_{ij} = 1 \mid \boldsymbol{y}_{ij}^{\mathsf{C}})}{(Y_{ij} = 0 \mid \boldsymbol{y}_{ij}^{\mathsf{C}})}\right] \tag{4}$$

$$= \boldsymbol{\theta}^{\mathsf{T}} \delta(\boldsymbol{y}_{ij}) \tag{5}$$

This means that the coefficients $\boldsymbol{\theta}$ are interpreted as the log-odds of an individual tie conditional on all other ties [37]. This is the major departure from the logit or probit model. The inclusion is necessary because $Pr(Y_{ij})$ is dependent on the dyad-wise outcome of all other dyads [23].